



Fotis Jannidis / Hubertus Kohle /
Malte Rehbein (Hg.)

Digital Humanities

Eine Einführung



J.B. METZLER



J.B. METZLER

Fotis Jannidis / Hubertus Kohle / Malte Rehbein (Hg.)

Digital Humanities

Eine Einführung

Mit Abbildungen und Grafiken

J.B. Metzler Verlag

Die Herausgeber

Fotis Jannidis ist Professor für Computerphilologie und Neuere Deutsche Literaturgeschichte an der Julius-Maximilians-Universität Würzburg.

Hubertus Kohle ist Professor für Mittlere und Neuere Kunstgeschichte an der Ludwig-Maximilians-Universität München.

Malte Rehbein ist Professor für Digital Humanities an der Universität Passau.

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

ISBN 978-3-476-02622-4

ISBN 978-3-476-05446-3 (eBook)

DOI 10.1007/978-3-476-05446-3

Dieses Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

J.B. Metzler, Stuttgart

© Springer-Verlag GmbH Deutschland, 2017

Einbandgestaltung: Finken & Bumiller, Stuttgart (Foto: shutterstock)

Satz: primustype Hurler GmbH, Notzingen

J.B. Metzler ist Teil von Springer Nature

Die eingetragene Gesellschaft ist Springer-Verlag GmbH Deutschland

www.metzlerverlag.de

info@metzlerverlag.de

Inhaltsverzeichnis

Warum ein Lehrbuch für Digital Humanities?.....	XI
---	----

I Grundlagen

1 Geschichte der Digital Humanities.....	3
1.1 Texte und Informationstechnologie: Der Gründungsmythos der Digital Humanities	3
1.2 Eine Community entsteht: Die frühen Jahre	5
1.3 Die Welt wird einfacher – Programmpakete.....	6
1.4 Die Welt wird noch einfacher – der Personal Computer	7
1.5 Vernetzungen von Personen und Ressourcen	9
1.6 Das WWW als einheitliches Interface	10
1.7 Das Beste kommt erst noch	11
2 Digital Humanities als Wissenschaft.....	13
2.1 Die Digital Humanities: Ein weites Feld	13
2.2 Die Digital Humanities: Werkzeug oder Methode?.....	13
2.3 Die Digital Humanities im Kontext der geisteswissenschaftlichen Disziplinen	14
2.4 Die Digital Humanities jenseits einzelner geisteswissenschaftlicher Disziplinen	15
2.5 Die Digital Humanities und die Informatik.....	17
2.6 »Die Digital Humanities«.....	18
3 Theorien digitaler Medien	19
3.1 Digitalisierung als Medien- und Wissensgeschichte	19
3.2 Medienarchäologie und Software Studies	22
3.3 Digitalisierung und Gesellschaft	26
3.4 Digitale Methoden.....	29
4 Aufbau des Computers und Vernetzung	35
4.1 Aufbau eines Computers	35
4.2 Eingabeperipherie, Ausgabeperipherie, Speicherperipherie	39
4.3 Computertypen.....	42
4.4 Benutzerschnittstellen	43
4.5 Vernetzung	45
4.6 OSI-Modell.....	46
4.7 TCP-IP-Modell	48
4.8 Netztopologie.....	50
4.9 Internet Protokoll.....	52
4.10 E-Mail und Webserver.....	54
5 Zahlen und Zeichen.....	59
5.1 Analog, digital und das Bit	59
5.2 Binäre Zahlen und Dezimalzahlen	60
5.3 Zeichenkodierung	61
5.4 Offene Probleme.....	66

6	Grundbegriffe des Programmierens	68
6.1	Anweisungen	69
6.2	Datentypen	70
6.3	Datenstrukturen 1: Listen	73
6.4	Ausdrücke, die Wahrheitswerte zurückgeben	74
6.5	Schleifen	77
6.6	Bedingte Verzweigungen	80
6.7	Datenstrukturen 2: assoziatives Feld	82
6.8	Modularisierung	84
6.9	Algorithmisches Denken	89
II	Datenmodellierung	
7	Grundlagen der Datenmodellierung	99
7.1	Grundbegriffe der Datenmodellierung	100
7.2	Stufen der Datenmodellierung	102
7.3	Datenmodellierung in der Praxis	104
7.4	Datenmodellierung in den Digital Humanities	106
8	Datenbanken	109
8.1	Datenverarbeitung und -organisation	109
8.2	Erstellen eines Datenmodells: Relationale Datenbank	112
8.3	Datenbankabfragen	123
8.4	Andere Datenbankmodelle	126
9	XML	128
9.1	Anwendung und Grundbegriffe	128
9.2	Grundstrukturen	130
9.3	Konzepte und Datenmodell	132
9.4	Modelle und Schemata	135
9.5	XPath	138
9.6	XSLT	141
9.7	X-Technologien im Einsatz	144
10	Netzwerke	147
10.1	Grundlagen	147
10.2	Rechnen mit Graphen	151
10.3	Den kürzesten Weg finden	154
10.4	Angewandte Netzwerkanalyse	157
10.5	Fazit	160
11	Ontologien	162
11.1	Begriff und Einordnung	162
11.2	Grundlegende Konzepte	164
11.3	Ein Beispiel	166
11.4	RDF	168
11.5	Speicherung, Retrieval und Datenintegration	172
11.6	Ontologien in den Digital Humanities	174

III Digitale Objekte

12	Digitalisierung	179
12.1	Grundlagen digitaler Bilder	181
12.2	Bilddigitalisierung	187
12.3	Erschließung der Digitalisate	192
12.4	Textdigitalisierung	193
12.5	Weitere Digitalisierungsverfahren	196
13	Digitales Publizieren	199
13.1	Eine Revolution	199
13.2	Neue Medien imitieren alte Medien	200
13.3	Eigenschaften der digitalen Publikation	201
13.4	Open Access	202
13.5	Neue Publikationsmodelle im Digitalen	204
14	Digitale Wissensproduktion	206
14.1	Was ist digitale Wissensproduktion?	206
14.2	Umgang mit Datenbanken	207
14.3	Crowdsourcing	208
14.4	Kollaboratives Schreiben: Wikipedia	209
14.5	Konsequenzen	211
15	Bibliothek, Archiv, Museum	213
15.1	Gedächtnisinstitutionen	213
15.2	Einheitliche Beschreibung von Objekten und Sammlungen	215
15.3	Wichtige Informationsportale	217
15.4	Gedächtnisinstitutionen als Forschungs- und Informationsinfrastrukturen für die Digital Humanities	220
16	Aufbau von Datensammlungen	223
16.1	Einleitung: Was sind Datensammlungen?	223
16.2	Erheben von Informationen über den gesamten Gegenstandsbereich ...	225
16.3	Sammeln, Zusammenführen und Säubern von Datensätzen	227
16.4	Erheben und Hinzufügen von Informationen über die Daten	228
16.5	Verfügbarmachen der Datensammlung	230
16.6	Fazit	232
17	Digitale Edition	234
17.1	Drei Beispiele zur Einführung	234
17.2	Worum geht es?	237
17.3	Definitiverischer Rahmen	238
17.4	Paradigmen digitaler Editionen	240
17.5	Methoden und die Realisierung digitaler Editionen	241
17.6	Editionen als Projekte	244
17.7	Technologien und Standards	245
17.8	Fragestellungen	248

IV Digitale Methoden

18	Manuelle und automatische Annotation	253
18.1	Was sind Annotationen?	253
18.2	Formalisierung und Operationalisierung	255
18.3	Annotationstypen und -funktionen	257
18.4	Annotationsverfahren	260
18.5	Objekte	262
19	Information Retrieval	268
19.1	Messwerte für IR-Systeme	269
19.2	Indexierung	270
19.3	Suchstrategien	271
19.4	Weitere Retrieval-Systeme	278
20	Quantitative Analyse	279
20.1	Was ist quantitative Analyse	279
20.2	Statistische Grundlagen	280
20.3	Maschinelles Lernen	289
20.4	Neuere Entwicklungen	297
21	Geographische Informationssysteme	299
21.1	GIS Datenmodelle	300
21.2	Raumprojektionen in gängigen Koordinatensystemen im Vergleich	303
21.3	Unsicherheiten von Raum- und Zeitangaben in historischen Quellen ...	305
21.4	Datenintegration und Kartentypen	307
21.5	GIS-Aufbau und -Komponenten	309
21.6	GIS-Methoden	311
21.7	Kartenrepositorien via Web WMS (World Map Services)	312
22	Digitale Rekonstruktion und Simulation	315
22.1	Definitionen	315
22.2	Digitale Rekonstruktion	315
22.3	Digitale Simulation	323
23	Informationsvisualisierung	328
23.1	Informationsvisualisierung in historischer Perspektive	328
23.2	Definitionen und Funktionen	330
23.3	Informationsvisualisierung im Forschungsprozess	331
23.4	Referenzmodell und Datenmodellierung	333
23.5	Datentypen	335
23.6	Visuelle Strukturen	336
23.7	Kritische Informationsvisualisierung	341

V Recht und Ethik

24	Recht	345
24.1	Digitale Objekte – mehr als nur ein Medienwandel.....	345
24.2	Digital Humanities und Open Access.....	350
24.3	Rechtsfragen digitaler Geistes- und Kulturwissenschaft: Ein Forschungsgegenstand der Digital Humanities?	351
25	Ethik	353
25.1	Ethische Fragen in den Digital Humanities: Eine Fallstudie	353
25.2	Moral, Ethik, Angewandte Ethik.....	354
25.3	Verantwortung als analytische Schlüsselkategorie.....	355

VI Anhang

26	Auswahlbibliographie	361
26.1	Allgemein.....	361
26.2	Fachspezifisch	361
26.3	Fachzeitschriften	362
	Autorinnen und Autoren	363
	Sach- und Personenregister	364

17 Digitale Edition

17.1 | Drei Beispiele zur Einführung

SCA Text Search Corpus Submit About Log in

Lectio 1, de Fide
By Petrus Plaoul

Edited by Jeffrey C. Witt Edition: 2011.10-dev-master | October 04, 2011
Original Publication: LombardPress, Sacramento, CA, March 23rd, 2011; License Availability: free,
Published under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 License

Lectio 1, de Fide
[Quaestio: utrum in causa iudiciali fidei contra traditionem pure humanitus
adinventam iudex idoneus ferret pro fide sententiam]

1 Circa prologum *Sententiarum* ^[1] in quo MAGISTER dicit quod intentionis suae est "munire Davidicam turrim
vel potius munitam ostendere clypeis ^[2] " ^[a] etc, quaero istam quaestionem: utrum in ^[3] causa iudiciali
fidei contra traditionem ^[4] pure humanitus adinventam iudex idoneus ferret pro fide sententiam. ^[5]

4 -- contra traditionem] contradictionem
R SV

Apparatus Fontium

- a -- LOMBARD, *Sentences*, I, prologus, (I:1)
- b -- Iacobus 1:17

Apparatus Criticus

- 1 -- Sententiarum] Magistri Petri Plaoul add R SV
- 2 -- clypeis] om. V
- 3 -- in] om. R SV S
- 4 -- contra traditionem] contradictionem R SV
- 5 -- quod] om. R V

Text Tool Bar Text Home Text Outline Previous Next Available Transcriptions Text Tools ▲

Max 50% Min Close

Base Text Comparison Text Generate

Base Text: critical
Circa prologum *Sententiarum* in quo Magister dicit quod intentionis suae est munire
Davidicam turrim vel potius munitam ostendere clypeis etc, quaero istam quaestionem:
utrum in causa iudiciali fidei contra traditionem pure humanitus adinventam iudex idoneus
ferret pro fide sententiam.

Comparison: reims
Circa prologum *Sententiarum* ~~magistri petri plaoul~~ in quo ~~Magister~~ dicit quod intentionis
suae est munire ~~Davidicam~~ turrim vel potius munitam ostendere clypeis etc. ~~Quaero~~
istam quaestionem: ~~utrum~~ in causa iudiciali fidei contra ~~traditionem~~ pure humanitus
adinventam ~~iudex~~ idoneus ferret pro fide sententiam.

Abb. 49 Petrus Plaoul: Sentenzenkommentar, ediert von Jeffrey C. Witt, Ohne Ort, 2011–2013. Neue Weboberfläche 2016. <petrusplaoul.org>. Screenshots mit oben = edierter Text, mitte = Apparat der Textbezüge und textkritischer bzw. Variantenapparat, unten = generierter Textvergleich

Diese Edition zielt auf einen **kritisch bearbeiteten »besten Text«**, der aus den überlieferten Dokumenten (Textzeugen) durch Vergleich (Kollation) und durch die Anwendung philologischer Kritik hergestellt wird. Während die vorhandenen Kopien der Texte (die »Handschriften«) relativ genau abgeschrieben (transkribiert) werden, ist der edierte Text gemäß der Grammatik und Orthografie des Lateinischen reguliert und mit einer neuen Interpunktion versehen. Die Edition bietet zwei Anmerkungsapparate: Nummern bezeichnen textliche Abweichungen bei den

Dokumenten (im Beispiel hat bei Stelle 4 der edierte Text ›contra traditionem‹, während die Handschriften aus Reims und St. Victor, die mit den Abkürzungen R bzw. SV bezeichnet werden, ›contradictionem‹ haben; in den Apparateinträgen steht ›add‹ = addidit/addiderunt für ›hat/haben hinzugefügt‹ und ›om‹ = omisit/omiserunt für ›hat/haben weggelassen‹); kleine Buchstaben bezeichnen Anmerkungen zu Bezügen zu anderen Texten. Die Edition bietet außerdem einen automatischen Vergleich zwischen dem edierten Text und den Fassungen der einzelnen Handschriften.

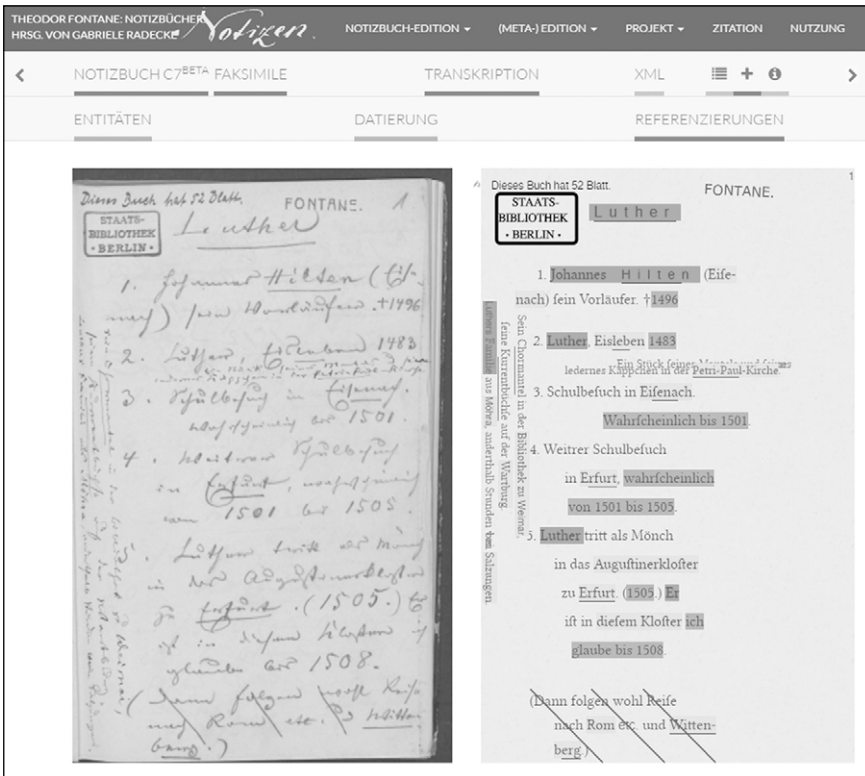


Abb. 50 Theodor Fontane: Notizbücher, Digitale genetisch-kritische und kommentierte Edition, herausgegeben von Gabriele Radecke, Göttingen 2015 ff. <fontane-nb.dariah.eu>. Screenshot zu Notizbuch C7

Die Fontane-Notizbücher werden sehr quellentreu und genau (sogar topografisch genau) abgeschrieben, um die zeitliche Entstehung des Textes (die **Textgenese**) verständlich zu machen. Dieses sehr genaue Vorgehen wird manchmal als ›diplomatish‹ bzw. ›hyperdiplomatish‹ Transkription bezeichnet. Die parallele Anzeige von Bild (›Faksimile‹) und Text nennt man **synoptische Darstellung**. Eine dritte hinschaltbare Darstellungsoption ist die Anzeige der zugrundeliegenden Daten im Format XML (s. Kap. 9). Der Text wird durch die Edition weiter **erschlossen**: Entitäten wie Orte oder Personen aber auch Datierungen und Verweise (Referenzierungen) werden ausgezeichnet, können farbig hervorgehoben (Personen = Magenta, Orte = Grün, Zeitangaben = Türkis) und in Registern zusammengeführt werden.

Jahrrechnungen der Stadt Basel 1535 bis 1610 – digital
Beta-Version

HOME Jahrrechnungen Projekt Erweiterte Suche Suche

Einnahmen
Einnahmen Stadt
Weinungeld
Mehlungeld
Stadtviehzoll
Bischofviehzoll
Pferdzoll
Torzölle
Wegzoll Neuer Weg
Wiesenbrückenroll
Gipszoll
Weinsticherbüchse
Wirtshausweinungeld
Kaufhauszoll
Pfundzoll
Hausgeld
Schultheissenstock Grossbas...
Lade
Gewinn Salzhandel
Schultheissenstock Kleinbasel
Brotkarren
Kornausfuhrzoll
Kaufhauswaage
Gerberzoll
Kembser Zoll
Diepflinger Zoll
Schiffzoll
Wiesenflösserzoll

☐ Alles in den Datenkorb übernehmen

fol. 1r [^]

Jarrechnung a festo
Johannis Baptistae
anno xv^c xxxv^o usque ad
festum Johannis Bap-
tistae anno xv^c xxxvj^o

fol. 1v [^]
fol. 2r [^]
fol. 2v [^]
fol. 3r [^]

☐ Empfangen

☐ Vom winungelt

☒ Prima angaria
☒ Secunda angaria
☒ Tercia angaria
☒ Quarta angaria

Suma iij^m vj^s lx lb

j^m lxxvij lb
vij^s box lb
vij^s lxxvij lb
ix^s xxvij lb

Jahrrechnung Stadt Basel
1535/1536

StaBS Finanz H 92.1

Rechnungslegungszeitraum:
26.6.1535 bis 24.6.1536

Transkription von:
Sonia Calvi
Lukas Meili

Zitiervorschlag: Jahrrechnung Stadt Basel
1535/1536, bearb. v. Sonia Calvi und Lukas
Meili, in: Die Jahrrechnungen der Stadt
Basel 1535-1610 - digitale Edition, hg. v.
Susanna Burghartz, Basel/Graz 2015
(zuletzt verändert am 2.3.2016), permalink
über DOI/handle: 11471/1010.1.1535

Datenkorb

Gesamtsumme:	3677,45 lb.
1535:	3677,45 lb.
Gesamtsumme:	3677,45 lb.

```
<pb facs="#fol_3r" xml:id="fol.3r"/>
<div ana="#bk_i #bs_Einnahmen" xml:id="bs_Einnahmen-div-0">
  <head xml:id="d2e145">Empfangen</head>
  <div ana="#bs_StadtEinnahmen" xml:id="bs_StadtEinnahmen-div-0">
    <div ana="#bs_Weinungeld" xml:id="bs_Weinungeld-div-0">
      <head xml:id="d2e152">Vom winungelt</head>
      <metamark function="aggregate" rend="Klammer"
        spanTo="#bs_Weinungeld-total-1" target="#bs_Weinungeld-1"/>
      <p ana="#bk_entry" rend="Klammer" xml:id="bs_Weinungeld-1">
        Prima angaria<seq ana="#bk_amount" rend="rb" xml:id="d2e159">
          <measure quantity="1083" type="currency" unit="lb">j<seq rend="super">xc<seq>lxxxiij lb</measure></seq>
```

Empfangen
Vom Winungelt
Prima Angaria
Secunda Angaria
Tercia Angaria
Quarta Angaria

Abb. 51 Jahrrechnungen der Stadt Basel 1535 bis 1610. Hg. von Susanna Burghartz, Basel/Graz 2015. DOI: 11471/1010.1.1535. Drei Screenshots: oben = Weboberfläche, unten links = XML-Code der Edition, unten rechts = Faksimile (jeweils Folio 3 recto)

Für die geschichtswissenschaftliche Forschung stehen die **Inhalte der Dokumente** im Vordergrund. Zugleich erschließt sich die Bedeutung der Texte und Zahlen aus der Struktur der Seite und der Einträge. Der Text wird deshalb nicht nur transkribiert, sondern in Form einzelner Einträge mit Werten und als Teile größerer Blöcke mit Summen erfasst. Die Rechnungsbereiche und Einzelposten können selektiert und in einem Datenkorb gesammelt werden. Verschiedene Export- und Download-Optionen (siehe die Icons oben rechts: tabellarische Darstellung, Faksimile, PDF, Basisdaten in TEI-XML, Daten im RDF-Format, Daten im CSV-Format) führen zu anderen Ansichten und ermöglichen die Weiterarbeit mit den Daten auch außerhalb der Edition.

17.2 | Worum geht es?

Wenn wir – im wissenschaftlichen Kontext, aber auch weit in die Alltagssprache hinein – von **Editionen** reden, dann sind fast immer ›kritische Editionen‹ gemeint: Wissenschaftlich aufbereitete Ausgaben von literarischen Texten oder historischen Dokumenten. Bei digitalen Editionen als einem der prominentesten Themen der Digital Humanities geht es dementsprechend um digitale wissenschaftliche Ausgaben, mit denen Grundlagenmaterial für die weitere Forschung bereitgestellt wird. Diese Forschung umfasst alle Bereiche und Disziplinen der Geistes- und Kulturwissenschaften und reicht von den verschiedenen Sprach- und Literaturwissenschaften (Bsp. 2: Neuere Deutsche Literaturwissenschaft), über die Geschichte (Bsp. 3: Wirtschafts- und Sozialgeschichte der frühen Neuzeit), die Kunstgeschichte bis hin zu den Medienwissenschaften, der Theologie oder der Philosophie (Bsp. 1: Theologie und Philosophie des Mittelalters).

Eine sehr alte wissenschaftliche Tätigkeit: Die Tradition der kritischen Auseinandersetzung mit den überlieferten Dokumenten und der Versuch, gute Fassungen und Ausgaben herzustellen, reichen sehr weit zurück. In der Kultur handschriftlicher Überlieferung bestand das Problem, dass originale Textfassungen verloren gingen, nur in Abschriften weitergegeben wurden und sich dadurch Fehler einschleichen konnten. Die Aufgabe der kritischen Philologie bzw. **Editionsphilologie** bestand deshalb darin, die Überlieferung zu sichten, die Abstammungsverhältnisse zwischen den Kopien zu ermitteln und aus den vorhandenen Varianten die jeweils besten zu ermitteln oder die ursprüngliche Textform einer bestimmten Stelle zu rekonstruieren um am Ende den **Urtext** (also jenen ersten, ursprünglichen Text, von dem letztlich alle Abschriften abstammen), den Autortext, den ›authentischen Text‹ oder zumindest einem dem Original möglichst nahe kommenden Text zu gewinnen und veröffentlichen zu können (s. Bsp. 1). Diese Methode der Sichtung der Überlieferung (Rezension), des Textvergleichs (Kollation), der Kritik der einzelnen Varianten (Examination) und ggf. der Textverbesserung (Emendation) gilt als editorische Methode der Textkritik und wird oft mit dem Namen Karl Lachmann verbunden verallgemeinert auch Lachmannsche Methode genannt, wenn sie mit dem Ziel der Konstruktion (**Konstitution**) eines idealen oder besten Textes verbunden ist. In den neu-philologischen Editionen ist man, da man häufig Autorhandschriften vorliegen hat, weniger an Überlieferungsvarianten als an Entstehungsvarianten interessiert, die einen Einblick in die Genese der Werke erlauben.

Dabei muss man beachten, dass die Herstellung eines Urtextes dort Sinn macht, wo die ältesten Fassungen verloren sind und Wert auf kanonische Fassungen gelegt wird, die sich auch stilistisch einem philologischen Ideal annähern. Dies ist vor allem für Texte aus der Antike oder dem Mittelalter der Fall. Neben der Erarbeitung eines »edierten Textes« kann es bei Editionen aber noch eine ganze Reihe weiterer **Zielstellungen** geben. Außerdem erstrecken sie sich auf ganz unterschiedliche **Materialarten, Dokumentsorten und Textgattungen**. Neben der »Lachmannschen« hat es immer schon andere Schulen und methodische Ansätze gegeben, die eine Konstruktion von Textfassungen jenseits der Wirklichkeit der Dokumente abgelehnt haben. Das kämpferische Schlagwort für diese letztlich historisch unbelegten und hypothetischen Editionstexte lautet dann: »*the text that never was*«.

Unterschiedliche Zielsetzungen: Jenseits der Rekonstruktion eines Urtextes, der Realisierung der Autorintention oder der Erfüllung des letzten Autorwillens (»Aus-

gabe letzter Hand«) kann es auch darum gehen, dem besten verfügbaren Zeugen zu folgen (Leithandschriftenprinzip), den tatsächlich rezipierten Text in den Vordergrund zu stellen, die Breite der Varianz sichtbar zu machen (Variorum-Edition), den Entstehungsprozess zu analysieren (genetische Edition; Bsp. 2), einen möglichst gut benutzbaren, »lesernahen« Text herzustellen (Bsp. 1 und 3), der Form der überlieferten Dokumente möglichst nahe zu kommen (diplomatische Ausgabe; s. Bsp. 2) oder vor allem den Informationsgehalt von Dokumenten herauszuarbeiten (Bsp. 3). Während der Schule der Textkonstitution und Textmischung aus den verschiedenen Textzeugen der Vorwurf des Eklektizismus gemacht worden ist, ist z. B. die amerikanische Editorik lange dem Prinzip des »copy text« (Leittext) gefolgt. Die französische Schule der »critique génétique« mit ihrer Konzentration auf die Genese literarischer Werke oder das amerikanische Konzept des »documentary editing« für die Erschließung historischer Dokumente ohne tiefere philologische Kritik sind weitere Beispiele für verschiedene methodische Strömungen.

Die **Zielstellungen** und die gewählte **Methode** hängen wesentlich vom jeweiligen Fachhintergrund und damit den Forschungsinteressen ab. So zielen Editionen aus den **Literaturwissenschaften** häufig auf einen philologisch bearbeiteten Text als Ausdruck eines sprachlichen Kunstwerks das z. B. früher orthografisch durchaus angepasst werden durfte (Bsp. 1). Dagegen muss es aus **sprachhistorischer Sicht** gerade um die Bewahrung schriftsprachlicher Eigenheiten gehen und die Edition näher an der spezifischen Form der Dokumente bleiben. Für die moderne Literaturwissenschaft sind auch die eigenhändigen Entwurfs- oder Notizhandschriften von Schriftstellern für die Erforschung des Schaffensprozesses besonders interessant – diese müssen dann ebenfalls so genau wie möglich – und sogar in ihrer zeitlichen Entstehungs- und Bearbeitungsfolge – wiedergegeben werden (Bsp. 2). Aus **historischer Sicht** wiederum bestehen zwei sich widersprechende Prinzipien: Auf der einen Seite geht es um eine historische Durchdringung und leichte Verständlichkeit des Inhalts, so dass Sacherkklärungen wichtig sind und der Text selbst lesernah normiert und (z. B. die Orthografie) gestaltet werden darf; auf der anderen Seite ist der Inhalt häufig nur in der Berücksichtigung der Form der Dokumente richtig zu verstehen, so dass hier detaillierte äußere Beschreibungen und sehr quellennahe Formen der Abschrift (Transkription) der Zeichen bis hin zur Berücksichtigung der zweidimensionalen Gestaltung (Layout, »mise en page«) der Dokumente wichtig sein können (Bsp. 3). Insgesamt hat in den letzten Jahrzehnten die Materialität der Dokumente, vom Zeichenträger bis zum Layout in den unterschiedlichen Disziplinen, die mit Editionen befasst sind, sehr zunehmende Aufmerksamkeit erfahren. Diese Tendenz ist durch die digitalen Editionen noch verstärkt worden.

17.3 | Definitivischer Rahmen

Editionen, und damit auch die digitalen Editionen, sollen **Grundlagen** für die Beantwortung geisteswissenschaftlicher Forschungsfragen bereitstellen. Sie befassen sich dazu mit den vorhandenen (überlieferten) Objekten, untersuchen sie, reichern sie mit **kritischem Wissen** an und führen schließlich zu Veröffentlichungen, die für einen breiten Nutzerkreis zugänglich sein sollen. Ein allgemeines Verständnis von Editionen muss alle historischen Zeiträume, alle theoretischen Ausgangspunkte, alle editorischen Schulen, alle Zielstellungen und alle zu edierenden Gegenstände abdecken. Dabei sind auch die folgenden beiden widerstreitenden Positionen zu verein-

nen: Für manche Forscher geht es bei der Edition um **Texte als abstrakte Objekte**, um Werke, die von physischen Objekten nur ›bezeugt‹ werden und die erst durch die Edition ›realisiert‹ werden. Andere sehen nur **Dokumente als reale Objekte** an, und die Editionen sind dann auch als Stellvertreter und erneute Medialisierungen abstrakter Texte wiederum kulturelle Artefakte. Da beide Seiten für sich eine ›realistische‹ Position beanspruchen, kann der Unterschied am ehesten durch die Begriffe ›Platonismus‹ versus ›Materialismus‹ gekennzeichnet werden. Im Spannungsfeld dieser Sichtweisen entfalten sich die möglichen Gegenstände der Edition. Sie umfassen alle schrifttragenden Objekte (also z. B. auch Inschriften) oder die Notation musikalischer Werke. Grundsätzlich ist aber alles edierbar, was einer kritischen Aufbereitung bedarf. Theoretisch also z. B. auch Bildwerke für die kunsthistorische Forschung oder ganz allgemein museale physische Objekte der Kulturgeschichte.

Allgemeine Definition

Edition ist die erschließende Wiedergabe historischer Dokumente. Dabei meint »historisch« eine zu überbrückende Distanz zwischen unserem heutigen Verständnis und dem Erklärungsbedarf zur historischen Wirklichkeit. »Dokumente« ist ein Oberbegriff für beschreibbare physische Gegenstände *und* die Texte, die mit ihnen kommuniziert werden. »Wiedergabe« bezeichnet das notwendige Kriterium der vollständigen Repräsentation des Editionsgegenstandes (merke: Eine Beschreibung, ein Katalog, ist noch keine Edition!) auf bildlicher und/oder (meistens) textlicher Ebene. »Erschließung« schließlich ist ein Oberbegriff für alle kritischen Operationen, von der Auswahl der zu edierenden Objekte über die äußere Beschreibung, die selbst in einer Transkription unvermeidlichen Entscheidungen, die Identifikation von referenzierten »Dingen« in Texten, die (historischen) Sachanmerkungen bis hin zur Textkritik im engeren Sinne (merke: Eine Reproduktion ohne Kritik ist noch keine Edition!).

Während hiermit die Edition im Allgemeinen beschrieben ist, wird die digitale Edition dadurch bestimmt, dass sie gewissen ›**digitalen Paradigmen**‹ folgt. Im Ergebnis geht sie damit über die traditionelle Edition hinaus und kann z. B. nicht ohne wesentliche Verluste an Inhalt oder Funktionalität in eine traditionelle (gedruckte) Edition verwandelt werden. Deshalb gilt auch: Eine (retro-)digitalisierte (vormals gedruckte) Edition ist (noch) keine (echte) digitale Edition. Denn eine digitalisierte Druckausgabe muss doch von ihrer Methode, ihren Inhalten und ihren Nutzungsformen her in der Denkwelt des Buchdrucks gefangen bleiben.

Breite versus tiefe Erschließung: Digitale Editionen stehen methodisch vielen anderen Objekten der Digital Humanities nahe. Ihr Inhalt ist zunächst das Ergebnis von Digitalisierungsmaßnahmen und besteht maßgeblich aus elektronischen Texten, die weiter ausgezeichnet und annotiert werden. Dadurch entstehen Datensammlungen, wie sie auch in Textkorpora, digitalen Bibliotheken, digitalen Archiven oder digitalen Museen vorliegen (s. Kap. 15 und 16). Im Unterschied zu diesen steht aber bei Editionen nicht eine möglichst große Menge gleichartig erschlossener Objekte im Vordergrund, sondern die **kritische, tiefe Erschließung eines thematisch begrenzten Gegenstandes**, der inhaltlich eingeführt, unter Umständen mehrfach (durch Abbildung, Transkription, kritischen Editionstext) repräsentiert, intern und extern vernetzt und durch zusätzliche Informationen und Materialien kontextualisiert wird. Schließlich ist die digitale Edition ein Sonderfall für Digitales Publizieren, weil sie häufig zu maßgeschneiderten, individuellen, sehr komplexen Publikationen führt.

17.4 | Paradigmen digitaler Editionen

Traditionelle und digitale Editionen unterscheiden sich durch die allgemeinen Paradigmen, denen sie folgen. Dies sind methodische Leitvorstellungen, die letzten Endes auch von den jeweiligen technischen Rahmenbedingungen abhängen. Für die traditionelle Edition sind es die Möglichkeiten und Beschränkungen der Druckkultur, die z. B. über die Ökonomie des Platzes die Menge der Inhalte einer Edition begrenzen und die Fokussierung auf »die Seite« als Grundfolie auf der etwas präsentiert werden kann. Die digitalen Medien kennen diese Grenze kaum, und so ist die digitale Edition durch eine grundsätzliche **Offenheit** gekennzeichnet, die es nahelegt, immer weitere Kontexte in die Edition einzubeziehen und schließlich zu einem Problem der **Entgrenzung** führt. Dagegen müssen Editoren mit der genauen Definition des zu edierenden Gegenstandes und seiner Grenze teilweise aktiv umgehen.

Work in progress: Die Bedingungen der Drucktechnologie sind auch Ursache für eine Fokussierung auf die *eine*, typografisch gefasste, kanonisierbare Form des konstituierten Textes. Die digitalen Medien führen dagegen zu einer weitgehenden Öffnung der Edition in verschiedenster Hinsicht: Das gedruckte Buch markiert einen Endzustand, es friert einen Wissensstand ein. Die digitale Edition ist leicht veränderlich und kann eine potentiell anhaltende Auseinandersetzung dokumentieren. Die Edition wird vom **Produkt** zum **Prozess**, schafft damit aber auch neue, bisher ungelöste Herausforderungen, was die Nachvollziehbarkeit von Eingriffen und Änderungen oder die stabile Zitierung betrifft. Die gedruckte Edition verweist implizit auf andere Texte und Informationen. Paratexte, Variantenapparate, Sachanmerkungen und Register bilden ein typografisch basiertes Informationsnetz, das in digitalen Editionen über Links unmittelbar nutzbar gemacht wird. Dabei wird die innere und äußere **Vernetzung** von ihrer typografischen Gebundenheit (Seite, Zeile, Fußnotennummer etc.) gelöst, »logifiziert«, systematisiert und verdichtet. Der digitale Link bezieht sich nicht auf eine Position in einer Publikation, sondern auf ein modelliertes Datum in einer digitalen Wissensbasis. Die gedruckte Edition war aus ökonomischen Gründen tendenziell bildfeindlich und monotextuell. In der digitalen Edition ist die Untermauerung durch Faksimiles nicht die Ausnahme, sondern die Regel. Diese Form der **Multimedialität** erhöht die **Transparenz** der wissenschaftlichen Bearbeitung, fördert damit aber auch eine materialistischere Sicht auf Texte.

Damit haben technische Veränderungen letztlich einen grundlegenden Einfluss darauf, was wir eigentlich unter den zu edierenden Texten verstehen und wie wir mit ihnen umgehen. Technologien verändern unseren **Textbegriff**. Digitale Editionen erlauben **mehrfache Textwiedergaben** auf der **Skala zwischen Quellennähe und Benutzernähe**: von topografischen über diplomatische, korrigierte, normierte, modernisierte Fassungen bis hin zu emendierten, konstituierten Texten. Auch das digitale Faksimile erscheint dann als erste Form der Textwiedergabe. Am anderen Ende der Skala markieren auch eine Übersetzung oder die Auflösung eines Textes in seine semantischen Aussagepartikel (z. B. in Form einer Datenbank) nur jeweils weitere Verarbeitungsschritte.

Grundlage dieser vielfachen Textformen sind reichhaltige Daten, die sowohl die Befunde aus den Dokumenten genau verzeichnen, als auch Deutungen und Interpretationen aufnehmen. In der elektronischen Recodierung wird davon gesprochen, dass die vorgefundene **Form** in ihren abstrakten **Inhalt** übersetzt wird. Oder umge-

kehrt, dass der Inhalt um Informationen zur Form ergänzt wird. Denn letztlich gibt es in der Haltung zum Text diese zwei gegensätzlichen Positionen: Ich habe einen Text als Dokument, das ich als solches beschreibe und recodiere und in dem ich dann einen Text als Zeichenkette finde – oder ich habe einen Text als Zeichenkette, zu der ich die Informationen zum Text als Dokument hinzufüge. Um zwischen der Datenbasis und ihrer Publikation zu unterscheiden, kann man die Begriffe **Repräsentation** (in modellbasierten Daten) und **Präsentation** (in Medien) verwenden.

Digitale Publikationen können nach dem *single-source*-Prinzip aus einer Datenquelle synchron und diachron immer wieder anders generiert werden. Man kann deshalb sagen, dass der eigentliche Wandel in der Editorik nicht so sehr in einem Wechsel der Medien liegt, als in ihrer **Transmedialisierung**: Bei Editionen geht es nicht nur um ihre mediale Erscheinung, sondern vor allem um das als Daten gefasste akkumulierte Wissen und das ihnen zugrundeliegende Modell als Explikation der editorischen Methode.

17.5 | Methoden und die Realisierung digitaler Editionen

Methodisch sind digitale Editionen von einigen wenigen zentralen Geboten geleitet. Hier ist zunächst der **wissenschaftliche Anspruch** grundlegend. Digitale Editionen sind Grundlagenforschung, die das bestehende Wissen zu einem bestimmten Thema bündeln, auf die Überlieferung anwenden und dadurch eine Basis schaffen, auf der weitere Forschungsfragen aufsetzen können. Die Edition nimmt in der Konfrontation der Befunde aus den Dokumenten mit dem Wissen über ihre Entstehung und ihre Intentionen eine **kritische Haltung** ein und verspricht, alle Teile gleichermaßen genau zu bearbeiten um am Ende einen gleichmäßig erschlossenen und bearbeiteten Dokument- oder Textbestand der Forschung zur Verfügung stellen zu können. Damit hängt das oberste, klar zu benennende Gebot der Editorik zusammen, das ausnahmslos für jedes Vorhaben gilt.

Grundregeln der Editorik

In Kenntnis der editorischen Methodologie im jeweiligen Gegenstandsbereich sollen für eine Edition sinnvolle, eindeutige und zielführende Regeln des Vorgehens aufgestellt werden. Dabei ist wichtig, diese einzuhalten und sie vollständig zu dokumentieren.

Nur dadurch kann gewährleistet werden, dass die editorischen Prozesse der Befundaufnahme, der Textbearbeitung und editorischer Entscheidungen **transparent** werden und die Entstehung der Edition **nachvollziehbar** bleibt. Dabei ist auch wichtig zu dokumentieren, *wer* für welche editorischen Arbeiten und Entscheidungen verantwortlich ist. Wenn eine Edition das Ergebnis der Anwendung des besten Wissens über einen Gegenstand ist, dann geht es auch um das Vertrauen in Akteure als Träger dieses Wissens.

Jenseits der editorischen Grundvoraussetzung eines regelgeleiteten wissenschaftlichen kritischen Prozesses kann eine digitale Edition als spezielle Form eines digitalen Informationssystems betrachtet werden. Eines der bekanntesten Architekturmuster in diesem Bereich hilft auch dabei, die einzelnen Arbeitsbereiche digitaler Editionen besser zu verstehen: Das »Model View Controller«-Paradigma (MVC) be-

schreibt Systeme, bei denen drei Dinge grundsätzlich getrennt voneinander betrachtet werden, auch wenn sie erst im Zusammenspiel gut benutzbare inhaltsreiche Anwendungen ergeben. *Model* steht dabei sowohl für das **Datenmodell** als auch die **Daten** selbst; *View* betrifft die Präsentation, die **Publikation**, die Darstellung der Edition; *Controller* beschreibt die technische Schicht dazwischen, die **Transformation** der Inhalte in eine digitale Präsentation, die auch durch Interaktion der Anwendung mit dem Benutzer bzw. der Benutzerin gesteuert wird. Die folgenden Abschnitte zeigen an beispielhaften Fragen, wie diese Bereiche abzudecken sind, wenn z. B. eine neue Edition konzeptionell entwickelt werden soll.

Model – Daten und Datenmodelle

Editionsmodelle: Die Grundlage aller digitalen Editionen sind Modelle. Dabei müssen verschiedene Bereiche explizit modelliert werden, die in den traditionellen Editionsprojekten teilweise nur nach impliziten »so hat man es schon immer gemacht«-Methoden gehandhabt wurden. Die Bereiche und Herausforderungen sind hier aber grundsätzlich immer schon ähnlich gewesen:

1. **Die Edition insgesamt:** Was ist der Gegenstand, was soll Inhalt der Edition sein? Welche Fragestellungen sollen mit der Edition bearbeitbar werden? Welche Kontexte (Dokumente, Informationen) sollen mitgegeben werden? Wie verhalten sich die Teile der Edition zueinander? Welche Bezüge zu Ressourcen außerhalb der Edition gibt es?
2. **Die Erschließung der Inhalte der Edition:** Wie sollen die Dokumente beschrieben werden? Welcher Metadatenstandard ist zu verwenden?
3. **Die Wiedergabe der edierten Gegenstände:** Wie sollen die visuellen Aspekte der Dokumente repräsentiert werden? Was sind die Parameter einer Bilddigitalisierung? Wie sollen Texte transkribiert werden? Wie dokumentnah soll die Transkription erfolgen? Geht die Repräsentation vom abstrakten Text zu seiner materiellen Erscheinung oder vom dokumenthaften Befund zu seiner Deutung über? Sollen topografische, grafematische, orthografische Phänomene berücksichtigt werden? Ob und wie sollen Abkürzungen (stillschweigend oder kenntlich) aufgelöst oder Normierungen vorgenommen werden? Dies sind nur wenige Beispiele für den potentiell sehr umfangreichen Regelungsbedarf, wenn es um die Transkription geht. Grundsätzlich ist hier zu modellieren, welche Dimension der Textualität wie abgedeckt werden soll: Visualität – Materialität – Zeichenbestand – Sprachlicher Ausdruck – (Werk-/Text-)Strukturen – Bedeutung / Semantik.
4. **Die weitere Kritik und Annotation der edierten Gegenstände:** Soll es eine philologische Kommentierung oder Annotation geben? Soll emendierend in den Text eingegriffen werden? Soll es einen historischen oder Sachkommentar geben? Wie ist mit Varianz bei mehrfach überlieferten Texten umzugehen? Werden variante Fassungen zusammengeführt oder als einzelne Texte belassen? Sollen benannte Entitäten (Personen, Orte etc.) identifiziert werden? Soll eine Erschließung durch Schlagworte erfolgen?

Auf der Grundlage der Modellierung stellt die Edition Inhalte bereit. Diese lassen sich grob zwei Gruppen zuordnen:

1. **Repräsentationsformen der edierten Gegenstände oder anderer Materialien:** Digitale Abbildungen und Transkriptionen, die immer weiter ausgezeichnet, an-

gereichert und annotiert werden, können verschiedene Sichten auf die Dokumente abbilden und erlauben in der Publikation mehrfache Wiedergabeformen.

2. **Zusätzliche Informationen:** Editionen haben vielfältige Paratexte: Vorworte, Einleitungen und eine Dokumentation ihrer selbst; oft gibt es weitere erläuternde Texte, Diskussionen oder eingebundene Sekundärliteratur. Zusätzliche Informationen können auch biografische Dossiers sein. Aus den Transkriptionen und Erschließungsdaten werden eigenständige Register und Wissensbasen generiert. Viele Editionen verfügen über Bibliografien zu ihrem Forschungsgegenstand.

View – Präsentation und Publikation

Hybridpublikation: Editionen können auf mehrfache Weise bereitgestellt werden. Auch bei digitalen Editionen spricht nichts gegen eine »Hybrid-Ausgabe«, also eine doppelte Publikation in digitalen und analogen Medien. Eine zusätzliche gedruckte **Buchausgabe** ist dann sinnvoll, wenn ein entsprechendes Nutzungsszenario, z. B. eine sequentielle, kontemplative Rezeption, zu erwarten ist. Es ist aber klar, dass eine solche Ausgabe gegenüber dem Editionsprojekt insgesamt nur spin-off-Charakter haben kann. Auf der anderen Seite können Editionen einfach als **Daten** an einer Schnittstelle, an einer Programmierschnittstelle (API) oder als Bulk-Download an einer Adresse bereitgestellt werden – um eine Weiterverwendung in anderen Kontexten oder eine systematische Auswertung zu erlauben. Der Regelfall ist aber die Veröffentlichung als **digitale Publikation**, für die dann die gleichen Regeln gelten, wie für andere komplexe Informationssysteme auch. Dabei ist dann u. a. zu regeln,

- wie das Gesamtsystem aussehen soll, wie es alle Inhalte gleichermaßen einfach zugänglich macht und welche Funktionalitäten es umfassen soll
- wie übersichtliche Browsing-Zugänge zu den Inhalten aufzubauen sind
- welche Formen einer allgemeinen oder spezialisierten Suche es geben soll
- wie mittels Registern, Indizes oder visualisierenden Aufbereitungen zusätzliche Wege zum Material geschaffen werden können
- wie die Inhalte selbst präsentiert werden
- wie verschiedene Formen der Inhalte (z. B. Faksimile und Transkription) zusammen dargestellt werden
- wie die Vernetzung der Inhalte untereinander oder mit externen Ressourcen umgesetzt wird
- wie die Inhalte über dauerhafte (persistente) Adressen feingranular ansprechbar, referenzierbar und zitierbar gemacht werden

Controller – Die technische Schicht

Digitale Editionen können als Publikationen verstanden werden. Man kann darunter aber auch die Daten der Edition fassen. Die vermittelnde Schicht, die die Daten verwaltet und die Publikation generiert ist dagegen eine grundsätzlich austauschbare Komponente. Ihre Aufgabe kann über die Generierung statischer Präsentationsformen hinausgehen, wenn der Benutzer bzw. die Benutzerin z. B. über Suchformulare, über Einstellungsmöglichkeiten oder über die Auswahl von zu vergleichenden Fassungen die Rückmeldung und das Verhalten der Edition selber steuern kann. Streng genommen kann deshalb zwischen einer **Publikationskomponente** und einer **Con-**

troller-Komponente im engeren Sinne unterschieden werden, die Nutzeraktionen aufnimmt und regelt, welche Ansicht (»View«) generiert werden soll. Für die Veröffentlichung der Edition gibt es viele verschiedene technische Lösungen. Zu den typischen Szenarien gehören ...

- die Generierung von eher statischen Webseiten aus den zugrundeliegenden Daten, z. B. mittels XSLT
- die Präsentation von Editionen mittels Content Management Systemen
- die Verwendung von Datenbanken mit einer zusätzlichen Programmierschicht zur Abfrage und Generierung von Webseiten
- der Einsatz von XML-*publication frameworks*

In letzter Zeit hat sich bei vielen großen, anspruchsvollen Editionen eine Architektur etabliert, die davon ausgeht, dass die zugrundeliegenden XML-Daten in XML-Datenbanken (»eXist«, »BaseX«) oder anderen Repository-Systemen (z. B. »fedora«) verwaltet und mit einer eigens programmierten Abfragelogik angesprochen und in eine Web-Präsentation verwandelt werden. Welche Technologien in einer Edition für die Schicht zwischen den Daten und der Publikation eingesetzt werden, hängt vor allem von den Gegenständen und der Komplexität der Edition, von den Kompetenzen und Vorlieben der Verantwortlichen und vom finanziellen und zeitlichen Rahmen eines Projektes ab.

17.6 | Editionen als Projekte

Komplexitätsfragen: Digitale Editionen werden fast immer als private Vorhaben (z. B. im Rahmen einer Dissertation), als öffentlich geförderte Drittmittelprojekte oder als Teil der Grundaufgaben von bestimmten Institutionen (Forschungseinrichtungen, Bibliotheken, Archive) durchgeführt. Sie neigen dazu, noch komplexer zu werden als ihre traditionellen Vorgänger. Es ist deshalb oft wichtig, für die vielen verschiedenen Aufgaben und Arbeitsbereiche in einem Team **Spezialisten** zu versammeln, die unterschiedliche Rollen wahrnehmen. Während dies in großen, anspruchsvollen Projekten tatsächlich zu einem Cast führen kann, der dem Abspann eines Kinofilms nahekommmt, sind allein agierende Editoren oder Editorenteams in kleineren Projekten gezwungen, viele Rollen zu übernehmen und die unterschiedlichsten Aufgaben selbst durchzuführen.

Für die Erstellung von digitalen Editionen gibt es kaum umfassende, leicht einsetzbare spezielle **Werkzeuge**. Vielmehr arbeiten die meisten Projekte mit allgemeinen, nicht auf Editionen ausgerichteten Werkzeugen, einem selbst zusammengestellten Baukasten, eigens entwickelten Werkzeugen oder angepassten Werkzeugen, um z. B. Transkriptionen zu erstellen und weiter zu bearbeiten. Gerade im Bereich der **Transkriptionstools** gibt es zwar einerseits eine sehr große Fülle von Entwicklungen, es hat sich aber andererseits bisher keine bestimmte Lösung durchgesetzt, die für verschiedene Arten von Texten und Editionen und für verschiedene methodische Herangehensweisen und Zielstellungen ohne hohen Anpassungsaufwand in einer größeren Zahl von Editionsprojekten einsetzbar wäre.

17.7 | Technologien und Standards

Für die unterschiedlichen Bereiche digitaler Editionen sind verschiedene **Datenformate, Standards** und **Technologien** zu berücksichtigen. Für die Inhaltsdaten selbst, also für digitale Bilder und Texte, gilt zunächst das gleiche wie in den Kapiteln zu Digitalisierung, Bibliothek, Archiv, Museum oder Datensammlungen gesagt. Hinter die Standards ›guter Digitalisierung‹ können vor allem Editionen nicht zurückfallen. Weil Editionen meistens komplexe, stark informationsangereicherte Textdaten enthalten, basieren sie heute fast immer auf **XML**. Hier sind zwar auch lokale ›XML-Dialekte‹ zur Modellierung und Kodierung von Editionstexten anzutreffen, als gegenwärtiger Leitstandard spielen aber die Empfehlungen der ›**Text Encoding Initiative**‹ (**TEI**) eine entscheidende Rolle (s. u.). Im Bereich der allgemeinen Wissenscodierung kommen dann auch noch die Normdaten oder auch Thesauri ins Spiel, die verschiedenen Formaten (z. B. wiederum XML) und konzeptionellen Standards (wie SKOS – Simple Knowledge Organization System) folgen können.

Als Publikationen sind Digitale Editionen heute in der Regel **Webanwendungen**, die auf den gängigen Webstandards beruhen und als HTML-Webpages auf Stilinformationen in Cascading Style Sheets (CSS) beruhen und häufig Javascript benutzen, um bestimmte zusätzliche Funktionalitäten zu ermöglichen. Die Inhalte von Editionen werden häufig zusätzlich als Druckexport angeboten wobei meistens PDF als Ausgabeformat verwendet wird. Vorlagen für den Druck von Büchern werden dagegen über manchmal komplizierte Verarbeitungspipelines mit »TUSTEP«, »LaTeX« oder »Adobe InDesign« erzeugt. Wenn Editionsdaten an technischen Schnittstellen bereitgestellt werden, dann können hier außer dem ursprünglichen Format (in der Regel XML) auch Exportformate wie Dublin Core (DC), METS, RDF (Resource Description Framework) oder andere fachspezifische Standards eine Rolle spielen.

In der Transformation der Daten für eine Publikation werden heute für anspruchsvolle Editionsprojekte häufig XML-Datenbanken mit einer Abfrageschicht aus XQuery und ggf. weiteren Skripten (z. B. in PHP) eingesetzt. Es gibt daneben aber viele andere Architekturen, die auf unterschiedlichen Datenbanksystemen und dem Einsatz verschiedener Programmiersprachen beruhen. Ein allgemeiner Standard ist heute noch nicht zu erkennen.

Text Encoding Initiative (TEI)

Der Editionsstandard: Das Herzstück der Edition ist die **Modellierung ihres Gegenstandes**, in der Regel ein oder mehrere Texte bzw. Dokumente. Für digitale Editionen gibt es im Bereich der grundlegenden Codierung von Dokumenten und Texten mit den Richtlinien der Text Encoding Initiative einen Standard, der unverzichtbar ist. Die aktuelle Fassung, P5 (Proposal 5) ist eine XML-basierte Sprache mit über 500 Elementen. Dabei wird zum einen eine Grundstruktur für die Beschreibung aller möglichen Texttragenden Objekte festgelegt; zum anderen bieten die Guidelines etliche Kapitel zur Auszeichnung von:

- **Textsorten:** Dichtung, Aufführungstexte, Wörterbücher, Handschriftenbeschreibungen, Gesprochene Sprache
- **Textphänomenen:** Namen, Daten, Personen, Orte; Tabellen und Grafiken
- **wissenschaftlichen Anwendungen:** Sprachkorpora, Kritische Apparate

- **textanalytischen Zwecken:** Verknüpfung, Segmentierung, Parallelisierung; Merkmalsstrukturen

Grundsätzlich wird in **TEI-Dokumenten** zwischen dem eigentlichen Text und einem *Header* für die Titeldaten, Beschreibung und Dokumentation der digitalen Fassung unterschieden. Der Text selbst wird meistens als abstrakter Text verstanden, der in einem Element `<text>` enthalten ist und sich dann über z. B. eine Buchstruktur (front, body, back) und Buchteile (div = Teil) zu einzelnen Absätzen (p = paragraph) weiter entfaltet. Er kann aber auch als Transkription aufgenommen werden, die sich an der physisch-visuellen Form von Quelldokumenten orientiert und in einem Element `<sourceDoc>` enthalten ist, das über `<surface>` einzelne »Schreiboberflächen« (z. B. Seiten) und darauf über `<zone>` einzelne »Schriftbereiche« bestimmt, die Text enthalten. Sieht man davon ab, dass selten eine abstrakte und eine materialistische Textsicht zusammen gegeben werden, dann sieht die Grundstruktur von TEI-Dokumenten ungefähr wie folgt aus, wobei die zunehmende Einrückung die formale Hierarchie von XML-Elementen andeutet:

TEI	
teiHeader	Alle Angaben außer dem Text selbst
fileDesc	Angaben zur vorliegenden Datei
titleStmt	Alle Titelinformationen
title	Titel des Dokuments
publicationStmt	
p	Informationen zur Publikation
sourceDesc	
p	Beschreibung der Quelle oder Vorlage
sourceDoc	Das transkribierte Dokument
surface	Eine beschriebene Oberfläche
zone	Ein Schriftraum
line	Eine Zeile mit Schrift
text	Ein (abstrakter) Text
body	Der Text selbst (ohne Paratexte)
div	Ein Textteil (z.B. ein Kapitel)
p	Ein Absatz

Nach den Regeln der TEI ausgezeichnete Texte folgen dem Prinzip von Auszeichnungssprachen im Allgemeinen und XML im Besonderen, nach denen Informationen einerseits in hierarchisch strukturierten Elementen organisiert werden und andererseits ein fortlaufender Text (Zeichenbestand) durch Elemente und Attribute annotiert wird. Die TEI realisiert die Idee, dass Texte **Ordered Hierarchies of Content Objects** (OHCOs) sind, also eine Abfolge von Inhaltsobjekten, die Teil einer

Hierarchie sind. Zugleich folgt sie dem Ansatz der **Übersetzung von Form in Inhalt**. Die visuelle Form von Texten repräsentiert dabei abstrakte Textstruktureinheiten, die im Code explizit gemacht werden können. Abgesetzte Blöcke sind dann Paragraphen (`< p >`) und zentrierte, größer gedruckte, vertikal abgesetzte Zeilen können als Überschriften (`< head >` = headline) erkannt werden. Textabschnitte, die kursiv gesetzt sind, sind dann herausgehobene Passagen (`< hi >` = highlighted) oder in einem »Nachdrücklichkeitsmodus« (`< emph >` = emphasized). Auf diese Interpretation kann aber auch verzichtet werden, wenn einfach ein Segment (`< seg >`) markiert wird, zu dem man festhält, wie es auf der Seite erscheint: `< seg rend = "italics" >` (`rend` = rendition). Auf der anderen Seite kann die Deutung bis zur **semantischen Explikation** vorangetrieben werden: `< persName key = "gnd_118580914" > Adolph Menzel < /persName >` – Die Zeichenkette »Adolph Menzel« steht für einen Personennamen der wiederum auf eine Person verweist, die in der Gemeinsamen Normdatei (GND) durch die Nummer 118580914 identifiziert wird.

Auf diese Weise unterstützt die TEI ein sehr breites Verständnis von Text, nach dem der digitale Text der Edition als Repräsentation texttragender Dokumente das Ergebnis einer filternden Wahrnehmung von Textphänomenen und der darauf aufbauenden Anwendung von regelgeleiteten Verarbeitungs-, Deutungs- und Annotationsprozessen ist. Mit der TEI können nicht nur **verschiedene Sichten auf Text**, also verschiedene Wahrnehmungs- und Interpretationsebenen codiert werden, sondern verschiedene **gleichzeitig**.

Begrenzungen von TEI: In seiner großen Fülle an Elementen und Attributen scheint die TEI auf den ersten Blick allumfassend und vollständig. Jeder Benutzer wird aber sehr schnell merken, dass für die eigenen, in der Regel sehr speziellen Fragestellungen die Haltung der TEI nicht ganz genau passend sein wird und gewünschte Elemente zur Beschreibung von Textphänomenen und zur Textannotation fehlen oder in bestimmten Kontexten nicht erlaubt sind. Die Richtlinien der TEI sind ein historisch gewachsenes Corpus, das methodische und fachliche Schieflagen hat. Manche Bereiche (Disziplinen, Textsorten, Betrachtungsweisen) sind besser ausgebaut als andere. Dies wird zum Teil dadurch gemildert, dass die TEI eine sehr große und aktive Community bildet, die beständig an der Weiterentwicklung der Richtlinien, des Element-Satzes und zusätzlichen Hilfsmitteln arbeitet. Zu der weiten **TEI-Infrastruktur** gehören neben der umfassenden Dokumentation verschiedene Gremien, Special Interest Groups, eine Mailingliste, eine jährliche Konferenz, eine eigene Zeitschrift und vielfältige Werkzeuge zur Erzeugung und Verarbeitung von TEI-Daten.

Wegen der Vielfalt der Gegenstände, der theoretischen Grundannahmen, der editorischen Schulen und der Zielstellungen ist für die Codierung von Texten in kritischen Editionen kein allgemeiner Standard im Sinne einer strikten Vorschrift möglich. Die TEI ist in den meisten Fällen nicht unmittelbar als Direktive für die editorische Arbeit anwendbar. Die vorgeschlagene Grundstruktur passt tatsächlich auf fast alle Anwendungsbereiche. Jenseits davon sind aber in der Regel **Anpassungen** nötig. Aus dem Gesamtset werden dazu meist lokal passende Elemente ausgewählt, Attributwertlisten festgelegt, Verwendungsregeln formuliert und zur Not auch neue Elemente geschaffen. Dieser Prozess führt dann auch technisch und formal zu lokalen XML-Schemata, für deren Erstellung die TEI mit ODD (One Document Does it All) eine Schema-Metasprache und mit »Roma« ein Werkzeug zur Entwicklung lokaler Schemata bereitstellt.

17.8 | Fragestellungen

Digitale Editionen als prominentes Anwendungs- und Forschungsgebiet der Digital Humanities befinden sich immer noch in der Entwicklung. Es bieten sich deshalb viele Fragestellungen für eine vertiefende Auseinandersetzung und für die Untersuchung aktueller Tendenzen an, die aus drei Bereichen kommen können.

1. Empirie

- Beschreibung, Evaluation und Kritik von einzelnen Editionen in ihren verschiedenen Aspekten (Faksimiles, Transkription, Textkonstitution, Kommentierung, Such- und Browse-Zugänge, Register, Paratexte etc.)
- Diskussion der Datenmodelle (z. B. des jeweiligen XML-Schemas oder der TEI-Regeln) für einzelne Editionen
- vergleichende Betrachtung der Ästhetik und Funktionalität digitaler Editionen: entwickelt sich ein neues Standardmodell der Publikation?
- Editionen als Daten, die an Schnittstellen bereitgestellt werden
- Beschreibung und Evaluation von Werkzeugen zur Transkription und Editions-erstellung

2. Praxis, Simulation

- Hypothetische Konzeption eines Editionsprojekts auf der Basis eines bestimmten Werkes oder Materialbestandes
- Entwicklung eines angepassten TEI-Modells für vorliegende Dokumente bzw. Texte
- TEI-Codierung exemplarischer Dokumente

3. Theoriebildung, offene Fragestellungen

- Mögliche Gegenstände der Editorik und dafür geeignete Methoden
- Adressierbarkeit und Zitation von Editionen und ihrer Inhalte auf einer feingranularen Ebene
- Rolle der Editoren und anderer Akteure in der Erarbeitung von Editionen
- Öffnung von Editionen für ihr Publikum als aktiver Teil des Editionsprozesses; Crowdsourcing.
- Umgang mit Textvarianz; was wird aus den kritischen Apparaten?
- Begriff und Konzept der Edition im Auflösen der Grenzen zu Digitales Archiv oder Virtuelle Forschungsumgebung

Literatur

- Apollon, Daniel/ B  lisle, Claire/R  gnier, Philippe (Hg.): *Digital Critical Editions*. Urbana/Chicaco/ Springfield 2014.
- Pierazzo, Elena: *Digital Scholarly Editing. Theories, Models and Methods*. Farnham 2015.
- Plachta, Bodo: *Editionswissenschaft – Eine Einf  hrung in Methode und Praxis der Edition neuerer Texte*. Stuttgart 1997.
- Sahle, Patrick: *Digitale Editionsformen*. Norderstedt 2013. Online: [http://kups.ub.uni-koeln.de/5351/ bzw. 5352, bzw. 5353](http://kups.ub.uni-koeln.de/5351/bzw.5352,bzw.5353) (31.10.2016).
- Sahle, Patrick: »Kriterien f  r die Besprechung digitaler Editionen« (2014), <http://www.i-d-e.de/publikationen/weitereschriften/kriterien-version-1-0/> (31.10.2016).

Internetquellen

A Catalog of Digital Scholarly Editions: www.digitale-edition.de. Eher retrospektive Sammlung von erschienenen »Editionen«, nicht von aktuellen Editionsprojekten.

Lexicon of Scholarly Editing: <http://uahost.uantwerpen.be/lse/> (2012 f.).

ride – A Review Journal for Digital Editions and Resources. Hg. vom Institut für Dokumentologie und Editorik: <http://ride.i-d-e.de>.

Text Encoding Initiative: <http://www.tei-c.org>.

Patrick Sahle